What is claimed is:

1. A system for analyzing unstructured documents for conceptual relationships, comprising:

a histogram module determining a frequency of occurrences of concepts in a set of unstructured documents, each concept representing an element occurring in one or more of the unstructured documents;

a selection module selecting a subset of concepts out of the frequency of occurrences, grouping one or more concepts from the concepts subset, and assigning weights to one or more clusters of concepts for each group of concepts; and

a best fit module calculating a best fit approximation for each document indexed by each such group of concepts between the frequency of occurrences and the weighted cluster for each such concept grouped into the group of concepts.

2. A system according to Claim 1, further comprising:

an extraction module extracting features from each of the unstructured documents and normalizing the extracted features into the concepts.

3. A system according to Claim 2, further comprising:

a structured database storing the extracted features as uniquely identified records.

4. A system according to Claim 1, further comprising:

a visualization module visualizing the frequency of occurrences, comprising at least one of creating a histogram mapping the frequency of occurrences for each document in the unstructured documents set and creating a corpus graph mapping the frequency of occurrence for all such documents in the unstructured documents set.

5. A system according to Claim 1, further comprising:

a threshold comprising a median and edge conditions, each such concept in the concepts subset occurring within the edge conditions.

6. A system according to Claim 1, further comprising:

an inner product module determining, for each group of concepts, the best fit approximation as the inner product between the frequency of occurrences and the weighted cluster for each such concept in the group of concepts.

7. A system according to Claim 6, wherein the inner product $d_{cluster}$ is calculated according to the equation comprising:

$$d_{cluster} = \sum_{i \to n} doc_{term_i} \cdot cluster_{term_i}$$

where $doc_{concept}$ represents the frequency of occurrence for a given concept in the document and $cluster_{concept}$ represents the weight for a given cluster.

8. A system according to Claim 1, further comprising:

a control module iteratively re-determining the best fit approximation responsive to a change in the set of unstructured documents.

9. A method for analyzing unstructured documents for conceptual relationships, comprising:

determining a frequency of occurrences of concepts in a set of unstructured documents, each concept representing an element occurring in one or more of the unstructured documents;

selecting a subset of concepts out of the frequency of occurrences;

grouping one or more concepts from the concepts subset;

assigning weights to one or more clusters of concepts for each group of concepts; and

calculating a best fit approximation for each document indexed by each such group of concepts between the frequency of occurrences and the weighted cluster for each such concept grouped into the group of concepts.

10. A method according to Claim 9, further comprising:

extracting features from each of the unstructured documents; and

normalizing the extracted features into the concepts.

11. A method according to Claim 10, further comprising:

storing the extracted features as uniquely identified records in a structured database.

12. A method according to Claim 9, further comprising:

visualizing the frequency of occurrences, comprising at least one of:

creating a histogram mapping the frequency of occurrences for each document in the unstructured documents set; and

creating a corpus graph mapping the frequency of occurrence for all such documents in the unstructured documents set.

13. A method according to Claim 9, further comprising:

defining a threshold comprising a median and edge conditions, each such concept in the concepts subset occurring within the edge conditions.

14. A method according to Claim 9, further comprising:

for each group of concepts, determining the best fit approximation as the inner product between the frequency of occurrences and the weighted cluster for each such concept in the group of concepts.

15. A method according to Claim 14, wherein the inner product $d_{cluster}$ is calculated according to the equation comprising:

$$d_{cluster} = \sum_{i \to n} doc_{term_i} \cdot cluster_{term_i}$$

where $doc_{concept}$ represents the frequency of occurrence for a given concept in the document and $cluster_{concept}$ represents the weight for a given cluster.

16. A method according to Claim 9, further comprising:

iteratively re-determining the best fit approximation responsive to a change in the set of unstructured documents.

17. A computer-readable storage medium holding code for performing the method according to Claims 9, 10, 11, 12, 13, 14, 15, or 16.

18. A system for dynamically evaluating latent concepts in unstructured documents, comprising:

an extraction module extracting a multiplicity of concepts from a set of unstructured documents into a lexicon uniquely identifying each concept and a frequency of occurrence;

a frequency mapping module creating a frequency of occurrence representation for each documents set, the representation providing an ordered corpus of the frequencies of occurrence of each concept;

a concept selection module selecting a subset of concepts from the frequency of occurrence representation filtered against a minimal set of concepts each referenced in at least two documents with no document in the corpus being unreferenced;

a group generation module generating a group of weighted clusters of concepts selected from the concepts subset; and

a best fit module determining a matrix of best fit approximations for each document weighted against each group of weighted clusters of concepts.

19. A system according to Claim 18, further comprising:

a histogram module creating a histogram mapping the frequency of occurrence representation for each document in the documents set.

20. A system according to Claim 19, further comprising:

a data mining module mining the multiplicity of concepts from each document as at least one of a noun, noun phrase and tri-gram.

21. A system according to Claim 19, further comprising:

a normalizing module normalizing the multiplicity of concepts into a substantially uniform lexicon.

22. A system according to Claim 21, wherein the substantially uniform lexicon is in third normal form.

23. A system according to Claim 18, further comprising:

a corpus mapping module creating a corpus graph mapping the frequency of occurrence representation for all documents in the documents set.

24. A system according to Claim 18, further comprising:

a threshold module defining the pre-defined threshold as a median value and a set of edge conditions and choosing those concepts falling within the edge conditions as the concepts subset.

25. A system according to Claim 18, further comprising:

a cluster module naming one or more of the concepts within the concepts subset to a cluster and assigning a weight to each concept with each such cluster.

26. A system according to Claim 25, further comprising:

a group module grouping one or more of the clusters into each such group of weighted clusters of concepts.

27. A system according to Claim 18, further comprising:

a Euclidean module calculating a Euclidean distance between the frequency of occurrence for each document and a corresponding weighted cluster.

28. A system according to Claim 18, further comprising:

a iteration module removing select documents from the documents set and iteratively reevaluating the matrix of best fit approximations based on a revised frequency of occurrence representation and concepts subset.

29. A system according to Claim 18, further comprising:

a structured database storing the lexicon, the lexicon comprising a plurality of records each uniquely identifying one such concept and an associated frequency of occurrence.

30. A system according to Claim 29, wherein the structured database is an SQL database.

31. A method for dynamically evaluating latent concepts in unstructured documents, comprising:

extracting a multiplicity of concepts from a set of unstructured documents into a lexicon uniquely identifying each concept and a frequency of occurrence;

creating a frequency of occurrence representation for each documents set, the representation providing an ordered corpus of the frequencies of occurrence of each concept;

selecting a subset of concepts from the frequency of occurrence representation filtered against a minimal set of concepts each referenced in at least two documents with no document in the corpus being unreferenced;

generating a group of weighted clusters of concepts selected from the concepts subset; and

determining a matrix of best fit approximations for each document weighted against each group of weighted clusters of concepts.

32. A method according to Claim 31, further comprising:

creating a histogram mapping the frequency of occurrence representation for each document in the documents set.

33. A method according to Claim 32, further comprising:

mining the multiplicity of concepts from each document as at least one of a noun, noun phrase and tri-gram.

34. A method according to Claim 32, further comprising:

normalizing the multiplicity of concepts into a substantially uniform lexicon.

35. A method according to Claim 34, wherein the substantially uniform lexicon is in third normal form.

36. A method according to Claim 31, further comprising:

creating a corpus graph mapping the frequency of occurrence representation for all documents in the documents set.

37. A method according to Claim 31, further comprising:

defining the pre-defined threshold as a median value and a set of edge conditions; and

choosing those concepts falling within the edge conditions as the concepts subset.

38. A method according to Claim 31, further comprising:

naming one or more of the concepts within the concepts subset to a cluster; and

assigning a weight to each concept with each such cluster.

39. A method according to Claim 38, further comprising:

grouping one or more of the clusters into each such group of weighted clusters of concepts.

40. A method according to Claim 31, further comprising:

calculating a Euclidean distance between the frequency of occurrence for each document and a corresponding weighted cluster.

41. A method according to Claim 31, further comprising:

removing select documents from the documents set; and

iteratively reevaluating the matrix of best fit approximations based on a revised frequency of occurrence representation and concepts subset.

42. A method according to Claim 31, further comprising:

storing the lexicon in a structured database, the lexicon comprising a plurality of records each uniquely identifying one such concept and an associated frequency of occurrence.

43. A method according to Claim 42, wherein the structured database is an SQL database.

44. A computer-readable storage medium holding code for performing the method according to Claims 31, 32, 33, 34, 36, 37, 38, 39, 40, 41, or 42.